III. DATA

CCMRP staff reviewed the clinical literature on pre-operative risk factors for bypass surgery and examined variables collected by the leading cardiac reporting programs to inform data collection for the program. CCMRP also reviewed a consensus statement prepared by a panel of researchers from the major CABG reporting programs that was particularly valuable in identifying those pre-operative characteristics of the patient that were most predictive of mortality (Jones et al., 1996). Appendix B contains a list of the variables identified in the consensus statement. Readers are directed to the ***California Report on Coronary Artery Bypass Graft Surgery: 1997-1998 Hospital Data Technical Report*** (July, 2001) for additional background on variable selection. Each year the data elements are reviewed and changes are made after consultation with the Technical Advisory Panel.

With some clarifications, CCMRP draws on a subset of data elements collected by the Society of Thoracic Surgeons (STS) for their National Database of Cardiac Surgery. Although the STS and CCMRP data definitions are virtually identical, CCMRP provides guidelines on interpretation of the definitions to assist hospitals with coding (see Appendix C). To improve the quality and comparability of data submitted across hospitals, CCMRP asks that each hospital receive training prior to beginning data submissions to CCMRP.

Table 3: CCMRP Data Elements, 1999*

1. Date of Surgery	2. Gender
3. Date of Birth	4. Race/Ethnicity (STS: Race)
5. Insurer (STS: Payor)	6. Patient's Zip Code
7. Height	8. Weight
9. Creatinine Level (Pre-operative)	10. Hypertension (Yes/No)
11. Dialysis (Yes/No)	12. Diabetes (Yes/No)
13. Peripheral Vascular Disease (Yes/No)	14. Cerebrovascular Disease (Yes/No)
15. Ventricular Arrhythmia (Yes/No)	16. Myocardial Infarction (MI) (Yes/No)
17. Date/Time of Most Recent MI (STS: MI When) (<=6 hrs., >6 but < 24 hrs., 1-7 days, 8-21 days, >21 days)	18. Number of Prior Cardiac Operations Requiring Cardiopulmonary Bypass
19. Date of Most Recent Cardiac Operation (STS: Previous CV Intervention: Most Recent)	20. Number of Prior PTCAs
21. PTCA/Atherectomy During Current Admission (STS: Prior PTCA including current admission)	22. PTCA to Surgery Time Interval (<=6hrs or >6hrs)
23. Chronic Obstructive Pulmonary Disease (Yes/No)	24. Congestive Heart Failure (Yes/No)
25. Angina (Yes/No)	26. Unstable Angina (Yes/No) (STS: Angina type: stable/unstable)
27. NYHA CHF Class	28. CCS Angina Class
29. Acuity (STS: Status) (elective, urgent, emergent, salvage)	30. Ejection Fraction (%)
31. Method of Measuring Ejection Fraction (LV Gram, radionuclide, or echocardiogram)	32. Left Main Stenosis (%)

33. Number of Diseased Vessels (None/Single/Double/Triple)	34. Mitral Insufficiency
35. Cross Clamp Time	36. Perfusion Time
37. Internal Mammary Artery (IMA) Used (Yes/No)	38. Cardioplegia (Yes/No)
39. Date of Discharge	40. Patient Status at Discharge (Alive/Dead)
41. Date of Death	

*See Appendix C for data element definitions

Hospital Data Submissions

Eighty-one hospitals initially submitted 21,973 usable records to CCMRP for the **1999 Analysis**. Sixty-eight of the 81 hospitals had previously submitted data for all or parts of 1997 and/or 1998.[12] As such, the combined rolled-up data across multiple years (1997-1999 **All Quarters** dataset) represents a total of 49,823 cases, with approximately 21% of the total cases from 1997 (10,391), 35% from 1998 (17,459), and 44% from 1999 (21,973).

The total number of cases submitted by each hospital varies across hospitals as a function of the size of the hospital and the date they commenced continuous participation in CCMRP. In other words, only records from hospitals that submitted continuously throughout the year with no submission "breaks" were included in the analyses. All hospitals shown in this report submitted data for a minimum of all four quarters of 1999. Appendix D presents a breakdown of each hospital's quarterly submissions.

Data Quality Review and Verification

CCMRP evaluated the data submitted from each hospital for completeness and potential data errors. The key steps involved in data cleaning and verification were:

- Step 1: Production and dissemination of hospital-specific data summary reports highlighting coding issues for clean-up;
- Step 2: Comparison of isolated CABG case volumes in CCMRP submissions with those in the OSHPD Patient Discharge Data (PDD);
- Step 3: Audit of a subset of cases at 36 hospitals and replacement of missing/inconsistent data with audited data;
- Step 4: CCMRP record linkage to the OSHPD PDD to evaluate accuracy of isolated CABG case submission and patient *Discharge Status* (alive/dead), with phone follow-up to hospitals to resolve resulting issues; and
- Step 5: Imputation of missing or invalid data values.

Hospitals that either refused audit (n=2) or had significant data problems that they were unable to fix (n=2) were dropped from the program.

Step 1: Hospital-Specific Data Summaries

Upon receipt of data at OSHPD, a "Quick Review Data Quality Check" form was filled out by the CCMRP Data Manager and immediately mailed to the hospital (see Appendix E). This one-page document noted potential problems based on a visual review of the distribution of data

[12] Enrollment in CCMRP is ongoing and hospitals can join at any time. Consequently, participants have varying numbers of quarters of data submissions, depending on the date they joined CCMRP.

element values in the dataset. Questionable cases were enumerated in a pre-printed list categorized into three problem types: missing data, logic problems, and out-of-range values. Hospitals were asked to immediately correct any problems noted and/or respond with additional explanatory information.

After receipt of all hospitals' 1999 full year data, hospitals were mailed a CCMRP Data Quality Report (DQR) (see Appendix F for item numbers 2 & 3 below). This report provides a detailed synopsis of the data received from each hospital, and compares each hospital's data submission with the aggregated data submitted by all hospitals during the time period. The DQR consisted of:

1. A cover letter explaining the report and its attachments.
2. A side-by-side univariate comparison (means and frequency distributions) of a given hospital's risk factors with those of all California hospitals submitting data to CCMRP.
3. A patient-level report detailing suspected errors based on data range checks, relational data edits, and missing critical data values.
4. A list of suspected duplicate records, when applicable.
5. A list of the hospital's inpatient deaths for the period.
6. Other pertinent program information, including definitions and imputation rules.

All hospitals received at least one DQR for the full 1999-year period, and most hospitals received more than one. Although the majority of hospitals made data corrections in response to coding issues identified in the DQRs, several hospitals did not respond, in spite of repeated requests by CCMRP staff.

Step 2: Comparison of Isolated CABG Cases: CCMRP vs. Patient Discharge Data
Corrections to CCMRP data based on the DQRs revealed hospital confusion concerning the CCMRP definition of an isolated CABG. A concern arose that hospitals were erroneously submitting non-isolated CABG cases to the program, and/or omitting cases from their CCMRP submission that were in fact isolated CABG surgeries. CCMRP's ability to evaluate this problem was limited by two factors: 1) the lack of a unique patient identifier with which to link CCMRP and PDD records (which represent all California hospital discharges) and, 2) the lack of an ICD-9-CM based definition of isolated CABG that could be employed to identify the target population in PDD.

As an interim step, staff compared each hospital's volume of isolated CABG cases as reported in the PDD (using a preliminary ICD-9-CM procedure code-based definition) with the number of cases submitted to CCMRP. This was done without linkage of records and without a formally tested and evaluated definition of isolated CABG based on ICD-9-CM codes.

Staff identified all hospitals with discrepancies between the two data sources that totaled more than 20 cases, or at least 10% of the hospital's volume. Thirteen hospitals met this criterion and were asked to explain the discrepancy. Ten of the thirteen hospitals discovered significant problems with their original submissions and subsequently made adjustments. In most cases, non-isolated CABG cases were eliminated from their submissions. Three hospitals maintained that no errors had occurred. Staff concluded that imprecision in the ICD-9-CM based definition of isolated CABG was likely responsible for these latter discrepancies.

Step 3: Data Replacements Using Medical Records Audit Data
Following preliminary data cleaning and analysis, CCMRP developed and implemented an audit process designed to formally review the quality of the data submitted for 1999. A subset of

cases at 36 of the 84 hospitals that originally submitted 1999 data were audited, representing 43% of hospitals submitting 1999 data and 12% of all usable records submitted to CCMRP.[13] The purpose and results of the audit are discussed briefly in the section below, **Audit of 1999 Data**, and in much greater detail in the ***California Report on Coronary Artery Bypass Graft Surgery 1999 Data, Technical Appendix: Audit Summary 2003*** (see www.oshpd.state.ca.us).

Auditor-abstracted data replaced the original hospital submission in the final model when data values recorded by auditors differed from those in the hospital's original submission. That is, for all data elements except *Discharge Status* (alive versus dead), the information submitted by the hospital was replaced by the information obtained in the audit. Our analysis of the audited data showed that this approach led to both improved risk model performance and improved data quality for audited hospitals.

The vast majority of audit data changes involved replacing missing values submitted by the hospitals with non-missing information obtained through the audit, though disagreement in coding of data elements at particular hospitals and across all hospitals was also noted. Additionally, the audit was used to verify that the cases selected for review were in fact isolated CABG surgeries. Audit results to the contrary (44 cases) were reviewed by CCMRP's medical consultant and, in all but five cases, resulted in removal of the record from the CCMRP analytic file.

Auditors sometimes had problems locating clear evidence of patient death in the medical charts alone. Findings of in-hospital death in OSHPD's PDD had proved highly reliable in previous studies, so PDD was considered the gold standard for recording patient deaths (Meux, 1990). This decision was validated by a subsequent CCMRP investigation into discrepancies in the coding of death among the PDD, CCMRP submission and audit findings at specific hospitals.

Step 4: Record-Specific Linkage of CCMRP Data with Patient Discharge Data Linkage
The audit revealed widespread problems with hospitals' coding of patient discharge status and interpretation of the definition of isolated CABG. CCMRP decided to conduct a linkage of the CCMRP dataset with the PDD in order to maximize the validity of the final results. Specifically, CCMRP records were linked, via a probabilistic matching algorithm[14], to all Patient Discharge Data records classified as Major Diagnostic Category 5 (MDC 5), Diseases and Disorders of the Circulatory System, as well as any records with ICD-9-CM code 36.1x in non-MDC 5 records. Also, an improved ICD-9-CM code-based definition of isolated CABG was developed to delineate those PDD records that could be isolated CABG surgeries.

CCMRP used this matched dataset to generate hospital reports when any of the three following conditions applied to patients whose *Discharge Status* was "dead" in either the PDD or CCMRP dataset:

1. There was a discrepancy in the discharge status of the patient between PDD and CCMRP (dead vs. alive).
2. An apparent isolated CABG mortality found in the hospital's PDD was not submitted to CCMRP (unreported death).
3. An apparent non-isolated CABG mortality was submitted to CCMRP (over-reported death).

[13] CCMRP audited all outlier hospitals identified at the time of the audit. During and subsequent to the audit, several hospitals either submitted data or replaced existing data with corrected information.

[14] A description of the methodology and mechanics of the data linkage are available from CCMRP upon request.

A total of 45 hospitals had cases meeting at least one of the above conditions. With regard to the first condition, CCMRP identified 17 cases in which patient *discharge status* was recorded as "dead" in the PDD, but reported as "alive" in the CCMRP submission. Alternatively, CCMRP also identified seven cases in which discharge status was recorded as "alive" in the PDD or the audit, but *discharge status* was recorded as "dead" in the CCMRP submission. The relevant hospitals were contacted and asked to review the specific cases. In all cases, the *discharge status* recorded in the PDD was found to be the correct information, and *discharge status* was appropriately re-coded.

For the second condition, 66 deaths from 32 hospitals were identified in the PDD as isolated CABG surgeries, but these cases were not found in the CCMRP submissions. In all cases, the hospital was contacted to explain the omission. Ultimately, 24 of the 66 records were confirmed as isolated CABG surgeries and submitted to CCMRP.

Regarding the third condition, eight deaths submitted to CCMRP from three hospitals could not be found in the PDD, or the PDD included ICD-9-CM codes suggesting that the cases were not isolated CABG surgeries. The hospitals were asked to review these cases and seven of the eight records were confirmed by the hospitals to be isolated CABG mortalities. The eighth record was found to have an incorrectly coded date of birth and was subsequently matched to its corresponding record in the PDD.

Step 5: Imputation of Missing or Invalid Data Values

Prior to running final risk models, it was necessary to impute missing or invalid data values so that all records could be retained in the model. When data were missing from the hospital submission, CCMRP replaced them with the lowest risk value for the variable in question. For example, if the hospital left the field *Diabetes* (Yes/No) blank, CCMRP assumed the condition was not present and assigned a "No" to that field. Likewise, if the value for the *NYHA congestive heart failure class* field was missing, we assigned the lowest risk category to this record—NYHA Class I.

The CCMRP policy decision to assign the lowest risk value to missing data was based on three factors: 1) many hospitals may leave data fields blank by design (e.g., blank means a co-morbid condition was not present or the value was normal); 2) consistency with the other major cardiac reporting programs, which replace missing data with the lowest risk or normal value; and 3) it creates an incentive for more complete coding by hospitals.[15]

In the case of the data element "creatinine," for example, the value was missing or recorded as "0" in approximately one-third of all cases submitted for analysis. In 1997, 1998, and 1999, the STS did not collect creatinine values unless those values exceeded 2.0. This coding practice made it impossible to distinguish between creatinine values below 2.0 (i.e., missing by design) and those that were truly missing (whether the value was below or above 2.0). Following the policy adopted for the 1997/1998 data collection, we assumed that all missing values of creatinine were "normal," and assigned them the value 1.0 mg/dl.

Between the 1997-98 and 1999 data collection periods, the percent of missing values decreased for most variables. In 1999, the variables with the largest number of missing values

[15] Note that in applying this policy, CCMRP replaced any missing values for the variable "coronary disease type" with the category found to be lowest risk in the All Quarters model: "double vessel disease." This rule differs from the one used for 1997-98 analysis in which missing values for this variable were replaced with the value "single vessel disease."

were: PTCA on same admission (N=8,513 or 38.3%), mitral insufficiency (N=7,835 or 35.2%), and left main stenosis (N=5,520 or 24.8%).

Hospitals with Unacceptable Quality Data

Not all hospitals responded to requests for data corrections and revisions with corrected data. Prior to producing the final risk-adjusted mortality results, staff and the Technical Advisory Panel had to decide whether data from any hospitals were so poor that their inclusion in the model would diminish overall predictive performance and lead to unreliable ratings for the hospitals in question. It was decided that data from two hospitals should be completely excluded from all analyses.

The internal data cleaning and external data validation processes used to generate this report appear to be more thorough than those used to produce similar statewide reports. CCMRP's efforts, however, were not exhaustive and have since been improved and expanded upon to ensure improved data integrity for reports in coming years.

Audit of 1999 Data

CCMRP developed and implemented an audit process designed to review the quality of the data submitted for 1999. Specifically, the 1999 data audit was designed to:

- Verify the accuracy of submitted data;
- Identify systematic coding problems that could compromise the validity of the statistical model;
- Determine if the rating received by a specific hospital was in any way a function of the hospital's coding practices. That is, did hospitals classified as better performers systematically overstate the severity of their cases (i.e., up-coding), or did hospitals classified as worse performers systematically understate the severity of their patient case-mix (i.e., down-coding); and,
- Determine the effect of CCMRP's policy to replace missing values with the lowest-risk category for each variable.

For 1999, 38 of 84 hospitals that originally submitted 1999 data were selected for audit, two of which refused audit and were dropped from the program. In total, 36 hospitals were audited, including all 16 hospitals that were identified in a preliminary analysis as either better or worse performers. An additional 20 hospitals were selected at random from the group of hospitals classified as "no different than expected."

Within each selected hospital, a subset of records was chosen for audit using a weighted random sample, in which all deaths were selected and records for more seriously ill patients were more likely to be selected. A total of 2,472 records, or 24% of all records submitted by the 36 targeted hospitals, were requested for audit. Overall, auditors were able to review 97.4% of requested records. The ***California Report on Coronary Artery Bypass Graft Surgery 1999 Data, Technical Appendix: Audit Summary 2003*** contains a detailed description of the audit process, analysis and findings. Summarized below are the analysis and key findings.

The audited data were compared against the data hospitals originally submitted to CCMRP. First, agreement statistics and bivariate frequencies were generated and analyzed in order to evaluate the accuracy of hospital coding for each variable. Second, a sensitivity analysis was

conducted to explore how hospital ratings would be affected if data submitted by the hospitals were replaced with the audit data.

The audit analysis found that most variables were coded acceptably, with the exception of *NYHA CHF Class*, *CCS Angina Class*, and *Acuity*. The poor coding of the *NYHA* and *CCS Class* variables had substantial implications for the validity of the model. Upon reviewing the audit findings, the CCMRP Technical Advisory Panel decided to exclude both *NYHA* and *CCS* class from the CCMRP model specifications. This decision had no effect on the fit of the final risk model or risk ratings of hospitals. The coding problems associated with *Acuity* were largely due to the subjective nature of coding this important variable. Hospitals identified as having severe coding problems with *Acuity* were asked to correct their data prior to the final analysis.

Other than the above-noted problems with *NYHA* and *CCS Class*, CCMRP did not find evidence of systematic coding problems among the hospitals classified as either "better" or "worse" than expected. However, hospitals rated "worse" than expected submitted, on average, more missing values in their data and tended to down-code (i.e., code as lower risk) more variables.

The decision to replace data originally submitted by the hospital with audit data led to several changes in hospital rankings (i.e., outlier status). The resulting changes were largely due to incorrect coding of the variables *Acuity* and *Discharge Status* (discussed earlier in this section). The audit data replacement strategy also resulted in significantly improved model performance.

All hospitals that were identified as outliers at the time the audit was conducted were audited. There were a few hospitals that submitted data after the audit commenced, one of which was identified as an outlier in the final analysis of the data; this single outlier hospital did not have its data audited because its data were received after the close of the audit.